

(12) 发明专利申请

(10) 申请公布号 CN 102902697 A

(43) 申请公布日 2013. 01. 30

(21) 申请号 201110226076. 2

(22) 申请日 2011. 07. 29

(71) 申请人 国际商业机器公司

地址 美国纽约阿芒克

(72) 发明人 姚毅 付国康 王冰欣 高雪峰

张大伟

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 吴立明

(51) Int. Cl.

G06F 17/30(2006. 01)

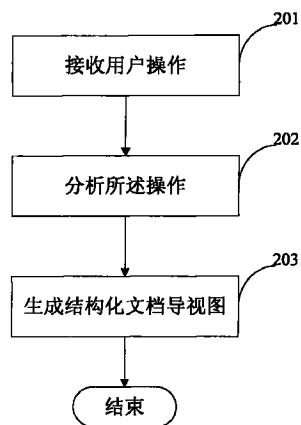
权利要求书 2 页 说明书 8 页 附图 4 页

(54) 发明名称

用于生成结构化文档导视图的方法和系统

(57) 摘要

本发明涉及文档处理技术领域。更具体地,本发明涉及用于生成可以体现用户的阅读兴趣和习惯的结构化文档导视图的方法和系统。本发明提供了一种用于生成结构化文档导视图的方法,包括:接收用户对所述文档的操作;对所接收到的操作进行分析;以及利用所述分析结果和所述文档的基础结构,生成结构化文档导视图。通过本发明,可以利用用户的阅读行为或所添加的注释来分析用户的阅读兴趣、习惯,并将分析结果映射为可直观地展示的结构化文档导视图,使得电子文档的读者能够直观地获取与用户的阅读兴趣、习惯相关的信息,以便利其阅读电子文档。



1. 一种用于生成结构化文档导视图的方法,包括:
接收用户对所述文档的操作;
对所接收到的操作进行分析;以及
利用所述分析结果和所述文档的基础结构,生成结构化文档导视图。
2. 如权利要求 1 所述的方法,其中对所接收到的操作进行分析包括:
响应于所接收到的操作是对所述文档的阅读行为,
记录与所述阅读行为相关的数据;并且
根据预定的规则对所记录的数据进行分析,其中所述规则定义了如何将用户的阅读行为对应至用户对所述文档的阅读兴趣或习惯。
3. 如权利要求 1 所述的方法,其中对所接收到的操作进行分析包括:
响应于所接收到的操作是在所述文档中输入注释,对所输入的注释进行语义处理,以体现用户对所述文档的阅读兴趣或习惯。
4. 如权利要求 2 所述的方法,其中与所述阅读行为相关的数据包括至少下列之一:在所述文档的某一部分停留的时间、在所述文档的若干部分之间的跳转以及对所述文档的某一部分的访问频率、用户对文档部分内容的外设操作。
5. 如权利要求 1-4 任一所述的方法,所述文档的基础结构是依据下列一种或多种信息获取的:所述文档的内容目录、所述文档的索引信息、所述文档的格式及样式信息。
6. 如权利要求 1-4 任一所述的方法,其中利用所述分析结果和所述文档的基础结构,生成结构化文档导视图包括:
生成所述文档的基础结构图;
设定代表不同含义的图例标记;以及
利用所述分析结果,将所述图例标记附加至所述基础结构图中。
7. 如权利要求 1-4 任一所述的方法,所生成的结构化文档导视图是可编辑的。
8. 一种用于生成结构化文档导视图的系统,包括:
接收装置,被配置为接收用户对所述文档的操作;
分析装置,被配置为对所接收到的操作进行分析;以及
导视图生成装置,被配置为利用所述分析结果和所述文档的基础结构,生成结构化文档导视图。
9. 如权利要求 8 所述的系统,所述分析装置被配置为:
响应于所接收到的操作是对所述文档的阅读行为,
记录与所述阅读行为相关的数据;并且
根据预定的规则对所记录的数据进行分析,其中所述规则定义了如何将用户的阅读行为对应至用户对所述文档的阅读兴趣或习惯。
10. 如权利要求 8 所述的系统,所述分析装置被配置为:
响应于所接收到的操作是在所述文档中输入注释,对所输入的注释进行语义处理,以体现用户对所述文档的阅读兴趣或习惯。
11. 如权利要求 8 所述的系统,其中与所述阅读行为相关的数据包括至少下列之一:在所述文档的某一部分停留的时间、在所述文档的若干部分之间的跳转以及对所述文档的某一部分的访问频率、用户对文档部分内容的外设操作。

12. 如权利要求 8-11 任一所述的系统,所述文档的基础结构是依据下列一种或多种信息获取的:所述文档的内容目录、所述文档的索引信息、所述文档的格式及样式信息。
13. 如权利要求 8-11 任一所述的系统,其中导视图生成装置被配置为:
生成所述文档的基础结构图;
设定代表不同含义的图例标记;以及
利用所述分析结果,将所述图例标记附加至所述基础结构图中。
14. 如权利要求 8-11 任一所述的系统,所生成的结构化文档导视图是可编辑的。

用于生成结构化文档导视图的方法和系统

技术领域

[0001] 本发明涉及文档处理技术领域。更具体地,本发明涉及用于生成可以体现用户的阅读兴趣和习惯的结构化文档导视图的方法和系统。

背景技术

[0002] 目前,电子阅读(e-Reading)技术已经得到越来越广泛的应用。相对于传统的纸质媒介而言,电子阅读以其便携、快速高效以及易于检索等特点受到了广大读者的欢迎。然而,在电子阅读中仍然存在很多不方便之处。

[0003] 例如,对于一本较长的电子书(或任何电子文档)而言,如果用户无法一次阅读完成,那么用户下次从中断部分继续阅读时可能需要重新阅读之前已经读过的部分重要章节(由用户所注释)以回忆起相关内容,但是电子书本身无法帮助用户判断哪些章节是重要章节,更无法快速定位到其所需要的重要章节。或者该用户在之前的阅读中经常在某些章节间来回切换(潜在地可能表明这些章节间有彼此间的关联),电子书同样无法以一种直观的方式帮助用户识别哪些章节间存在关联,或者存在何种类型的关联。用户多数情况下只能通过再次浏览电子书的内容而发现章节间的关联性、章节的重要性或找出该用户所添加过的注释。进一步地,对于该电子书或电子文档的其它读者而言,无法方便地借鉴在先读者通过其在先阅读行为或注释所体现出的对该电子书/文档的理解,例如哪些章节被长时间阅读、反复浏览,哪些章节间经常被直接跳转或切换,哪些章节被标注为重要性较高等等。

[0004] 很多电子文档中均包含类似于目录或索引的结构化信息,现有技术中存在利用这些文档中已有的结构化信息而生成结构化图形以展示文档结构的技术方案,但是这种结构化图形仅仅是对电子文档本身结构的展示,其中并不包含基于用户的阅读行为或注释而产生的体现用户阅读兴趣、习惯的信息(例如章节间的关联性、重要性等)。

发明内容

[0005] 考虑到上述存在的问题,本发明的目的之一在于提供一种能够通过用户对电子文档的阅读行为或所添加的注释来分析用户的阅读兴趣、习惯的技术方案。本发明的又一目的在于提供一种能够将分析结果映射为可直观地展示的结构化导视图的技术方案。本发明的再一目的在于提供一种利用已有的电子文档结构化信息,来直观地展示用户对电子文档的阅读兴趣、习惯的技术方案。本发明的又一目的在于提供一种能够让文档的用户借鉴、参考该用户阅读历史、兴趣、习惯的技术方案。本发明的再一目的在于提供一种能够让文档的用户借鉴、参考其它用户的阅读历史、兴趣、习惯的技术方案。

[0006] 根据本发明的一个方面,提供了一种用于生成结构化文档导视图的方法,包括:接收用户对所述文档的操作;对所接收到的操作进行分析;以及利用所述分析结果和所述文档的基础结构,生成结构化文档导视图。

[0007] 根据本发明的一个实施例,其中利用所述分析结果和所述文档的基础结构,生成

结构化文档导视图包括：生成所述文档的基础结构图；设定代表不同含义的图例标记；以及利用所述分析结果，将所述图例标记附加至所述基础结构图中。

[0008] 根据本发明的另一方面，提供了一种用于生成结构化文档导视图的系统，包括：接收装置，被配置为接收用户对所述文档的操作；分析装置，被配置为对所接收到的操作进行分析；以及导视图生成装置，被配置为利用所述分析结果和所述文档的基础结构，生成结构化文档导视图。

[0009] 根据本发明的方法和系统，可以通过用户的阅读行为或所添加的注释来分析用户的阅读兴趣、习惯，并将分析结果映射为可直观地展示的结构化文档导视图，使得电子文档的读者（可以是执行阅读行为的读者也可以是首次阅读该电子文档的读者）能够直观地获取与用户的阅读兴趣、习惯相关的信息，以便利其阅读电子文档。

附图说明

[0010] 通过对结合附图所示出的实施方式进行详细说明，本发明的上述以及其他特征将更加明显，本发明附图中相同的标号表示相同或相似的部件。在附图中，

[0011] 图 1 示出了适于用来实现本发明实施方式的示例性计算系统 100 的框图；

[0012] 图 2 示出了根据本发明一实施例的用于生成结构化文档导视图的方法的流程图；

[0013] 图 3 示出了根据本发明又一实施例的用于生成结构化文档导视图的方法的流程图；

[0014] 图 4 示出了根据本发明一实施例的所生成的结构化文档导视图的实例；

[0015] 图 5 示出了根据本发明一实施例的用于生成结构化文档导视图的系统的框图。

具体实施方式

[0016] 在下文中，将参考附图通过实施方式对本发明提供的用于的方法和系统进行详细地描述。

[0017] 所属技术领域的技术人员知道，本发明的多个方面可以体现为系统、方法或计算机程序产品。因此，本发明的多个方面可以具体实现为以下形式，即，可以是完全的硬件、完全的软件（包括固件、驻留软件、微代码等）、或者本文一般称为“电路”、“模块”或“系统”的软件部分与硬件部分的组合。此外，本发明的多个方面还可以采取体现在一个或多个计算机可读介质中的计算机程序产品的形式，该计算机可读介质中包含计算机可用的程序码。

[0018] 可以使用一个或多个计算机可读的介质的任何组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是——但不限于——电的、磁的、光的、电磁的、红外线的、或半导体的系统、装置、器件或任何以上的组合。计算机可读存储介质的更具体的例子（非穷举的列表）包括以下：有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任何合适的组合。在本文件的语境中，计算机可读存储介质可以是任何包含或存储程序的有形的介质，该程序被指令执行系统、装置或者器件使用或者与其结合使用。

[0019] 计算机可读的信号介质可包括在基带中或者作为载波一部分传播的、其中体现计

计算机可读的程序码的传播的数据信号。这种传播的信号可以采用多种形式,包括——但不限于——电磁信号、光信号或任何以上合适的组合。计算机可读的信号介质可以是并非为计算机可读存储介质、但是能发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序的任何计算机可读介质。计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括——但不限于——无线、电线、光缆、RF 等等,或者任何合适的上述组合。

[0020] 计算机可读介质上包含的程序代码可以用任何适当的介质传输,包括——但不限于——无线、电线、光缆、RF 等等,或者任何合适的上述组合。

[0021] 用于执行本发明的操作的计算机程序码,可以以一种或多种程序设计语言的任何组合来编写,所述程序设计语言包括面向对象的程序设计语言 - 诸如 Java、Smalltalk、C++ 之类,还包括常规的过程式程序设计语言 - 诸如“C”程序设计语言或类似的设计语言。程序码可以完全地在用户的计算机上执行、部分地在用户的计算机上执行、作为一个独立的软件包执行、部分在用户的计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在后一种情形中,远程计算机可以通过任何种类的网络——包括局域网 (LAN) 或广域网 (WAN) - 连接到用户的计算机,或者,可以(例如利用因特网服务提供商来通过因特网) 连接到外部计算机。

[0022] 以下参照按照本发明实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述本发明的多个方面。要明白的是,流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合,都可以由计算机程序指令实现。这些计算机程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器,从而生产出一种机器,使得通过计算机或其它可编程数据处理装置执行的这些指令,产生实现流程图和/或框图中的方框中规定的功能/操作的装置。

[0023] 也可以把这些计算机程序指令存储在能指令计算机或其它可编程数据处理装置以特定方式工作的计算机可读介质中,这样,存储在计算机可读介质中的指令产生一个包括实现流程图和/或框图中的方框中规定的功能/操作的指令装置(instruction means)的制品。

[0024] 也可以把计算机程序指令加载到计算机或其它可编程数据处理装置上,使得在计算机或其它可编程数据处理装置上执行一系列操作步骤,以产生计算机实现的过程,从而在计算机或其它可编程装置上执行的指令就提供实现流程图和/或框图中的方框中规定的功能/操作的过程。

[0025] 下面参看图 1。图 1 示出了适于用来实现本发明实施方式的示例性计算系统 100 的框图。如所示,计算机系统 100 可以包括:CPU(中央处理单元)101、RAM(随机存取存储器)102、ROM(只读存储器)103、系统总线 104、硬盘控制器 105、键盘控制器 106、串行接口控制器 107、并行接口控制器 108、显示控制器 109、硬盘 110、键盘 111、串行外部设备 112、并行外部设备 113 和显示器 114。在这些设备中,与系统总线 104 耦合的有 CPU 101、RAM 102、ROM 103、硬盘控制器 105、键盘控制器 106、串行控制器 107、并行控制器 108 和显示控制器 109。硬盘 110 与硬盘控制器 105 耦合,键盘 111 与键盘控制器 106 耦合,串行外部设备 112 与串行接口控制器 107 耦合,并行外部设备 113 与并行接口控制器 108 耦合,以及显示器 114 与显示控制器 109 耦合。应当理解,图 1 所述的结构框图仅仅为了示例的目的而

示出的,而不是对本发明范围的限制。在某些情况下,可以根据具体情况而增加或者减少某些设备。

[0026] 图 2 示出了根据本发明一实施例的用于生成结构化文档导视图的方法的流程图。图 2 所示的方法从步骤 201 开始。在步骤 201 中,接收用户对文档的操作。本发明中的“文档”,如无特别说明,均指“电子文档”。本发明中的“电子文档”包括所有以数据的形式在电子媒介中存储的文本、表格、可读指令、图像、甚至是音视频内容。根据本发明的一个实施例,“电子文档”中包含了结构化信息。所谓文档的结构化信息,是指电子文档中以数据形式存储的关于文档基础结构的信息,例如包括电子文档的各个部分之间的层级关系及其各个部分的简要信息(如章节代码、标题等)。以 Office Word(微软公司注册商标)软件为例,对于一篇 Word 文档而言,可以通过“大纲视图”的选项来显示其结构化信息,其中可以展示各个段落所属的层级以及各个段落间的从属关系。在编辑 Word 文档时也可以在“大纲视图”的选项下对段落间的层级关系进行设定。有的 Word 文档还具有目录及索引信息,这样就更直接地提供了电子文档的结构化信息。本领域技术人员应当理解,不同的电子文档处理软件可能具有不同的存储或表示其结构化信息的方式,诸如分段符(通常响应于“回车键”输入)、分节符等都可以用于记录或表示电子文档的结构化信息。即使在没有现成的目录或索引信息的情况下,也可以通过分析文档的格式和样式,如标题级别、段落换行、页码等来提炼出电子文档的结构化信息。

[0027] 根据本发明的一个实施例,用户对电子文档的操作包括用户对电子文档的阅读行为,例如用户在阅读过程中停留在电子文档中的某一部分(章节或段落)的时间、阅读某一部分的频率、在电子文档的若干部分之间的跳转关系等。这是基于如下的考虑:上述列举的阅读行为均可以从某种程度上反映用户在阅读电子文档过程中的阅读兴趣或习惯。例如,如果用户在电子文档中的某一段落停留了很长时间或多次停留在该段落,则说明用户需要很长时间来读懂该部分内容或者需要反复多次阅读该部分内容,从而进一步说明该部分内容比较重要或者比较难以理解。又如,如果用户经常在电子文档的第二段和第四段间跳转,则说明这两个段落间存在某种关联性,经常需要结合起来阅读。本领域技术人员应当理解,用户对电子文档的阅读行为可能包含很多其它的类型,绝不仅限于上述列举的几种。例如,用户直接删除电子文档中的某一段落也是一种阅读行为,该操作可以反映用户认为该段落没有价值等含义。又例如用户对文档部分内容的外设操作,如鼠标选择反显部分内容,打印某段内容等也均属于用户对文档的阅读行为。

[0028] 根据本发明的又一实施例,用户对电子文档的操作包括用户在电子文档中输入注释。所输入的注释既可以是文字的记录(例如在某一段落结尾处输入“精彩”),也可以是符号标识(例如在某一页结尾处输入“?”以表示读者对该页内容有疑惑)。所输入的注释体现了用户在阅读电子文档过程中的兴趣或者特定情绪。

[0029] 由于用户对电子文档的操作是通过存储或展示电子文档的电子媒介进行的,因此用户的操作自然可以被电子媒介所记录。

[0030] 接下来,图 2 所示的方法进行至步骤 202。在步骤 202 中,对所接收到的操作进行分析。从上文中可知,由于所接收到的用户对电子文档的操作体现了用户在阅读过程中的兴趣、习惯、关注点或者某些特定情绪,因此为了将这些阅读兴趣、习惯以可量化的方式直观展示在结构化文档导视图中,就需要对所接收到的操作进行统计分析。下文中将结合图

3 具体描述根据本发明多种实施例的分析方法。但是,本领域技术人员应当理解,无论采用何种方式、利用何种规则进行分析,只要分析结果试图反映(无论是定性地还是定量地)用户的阅读兴趣、习惯,即落入本发明保护范围。

[0031] 在步骤 203 中,利用分析结果和电子文档的基础结构,生成结构化文档导视图。“结构化文档导视图”说明:(1) 该图是结构化的或者以结构化的方式表征的图形,例如树形图、层次关系图、目录索引图等均为结构化的图;(2) 该图具有导视作用,这里所指的导视作用不仅仅是对文档目录、段落、分页等简单层次信息的导视,而是指包含了用户阅读兴趣、习惯等信息的结构化导视。可以理解,由于在步骤 202 中对用户操作的分析是结合了电子文档的基础结构信息进行的对用户阅读兴趣、习惯的分析,因此利用分析结果和电子文档的基础结构,就可以生成包含了用户阅读兴趣、习惯信息结构化文档导视图。下文中将结合图 3 和图 4 具体描述生成结构化文档导视图的实施例。

[0032] 综上所述,利用图 2 示出的方法,可以通过用户对电子文档的操作来分析用户的阅读兴趣、习惯,并将分析结果映射为可直观地展示的结构化文档导视图,使得电子文档的读者(可以是执行阅读行为的读者也可以是首次阅读该电子文档的读者)能够直观地获取与用户的阅读兴趣、习惯相关的信息,以便利其阅读电子文档。

[0033] 图 3 示出了根据本发明又一实施例的用于生成结构化文档导视图的方法的流程图。图 3 所示的方法从步骤 301 开始,在步骤 301 中,接收用户对电子文档的操作,步骤 301 对应于步骤 201,在此不再赘述。

[0034] 接下来,进行至步骤 302,判断所接收到的用户操作的类型。如果所接收到的用户操作为用户对电子文档的阅读行为(参见上文对

[0035] “阅读行为”的说明),则进行至步骤 303,记录与阅读行为相关的数据。根据上文中对“阅读行为”的说明,与阅读行为相关的数据包括例如用户在电子文档某一部分停留的时间长度、用户在单位时间内阅读电子文档某一部分的次数、电子文档的各个部分之间跳转的顺序等。接下来在步骤 304 中,根据预定的规则对所记录的数据进行分析,其中所述规则定义了如何将用户的阅读行为对应至用户对电子文档的阅读兴趣或习惯。例如,预定的规则可以是:If time > 60s, then mark the paragraph as “important”,这一规则的含义是“如果用户在电子文档的某一段停留的时间超过 60 秒,则该段属于重点段落”。预定的规则还可以是:If direct jump > 5, then mark the two paragraphs as “related”,这一规则的含义是“如果直接跳转次数大于 5,则这两个段落间具有相关性。还有一些预定的规则可能是基于比较的,例如“阅读次数最多的段落即为用户最感兴趣的段落”这一规则即需要对不同段落的阅读次数进行统计并最终得出分析结果。本领域技术人员应当理解,上面只是给出了三个关于预定的规则的示例,完全可以根据需要预定不同的规则用于分析用户阅读行为数据。

[0036] 下面回到步骤 302,如果在步骤 302 中判断出所接收到的用户操作为用户在电子文档中输入注释,则进行至步骤 305。在步骤 305 中,对所输入的注释进行语义处理,以体现用户对所述文档的阅读兴趣或习惯。例如,如果用户在电子文档的某一段结尾处输入问号“?”,则表明用户对该段存有疑惑。这里所说的语义处理即是指将用户输入的注释“?”进行语义处理,分析得到这个注释的语义——对该段存有疑惑。又例如,如果用户将文档中某一段的文字内容选中然后加以特殊颜色背景以区分与其它文字,则通过语义处理将特殊

颜色背景所覆盖的文字分析为——该内容比较重要。根据本发明的一个实施例,对用户输入的注释进行语义处理是根据预定的规则进行的。这里所指的预定的规则可以是例如:“? ——疑惑内容”(表示用户如果输入“?”,则表示对该段内容有疑惑)、“ifbackground color = blue, then the covered text is important”(表示如果部分文字内容为蓝色背景,则表示该部分为重要内容)。本领域技术人员应当理解,上面只是给出了两个针对用户输入注释的预定的规则的示例,完全可以根据需要预定不同的规则用于对用户输入的注释进行语义处理。

[0037] 图3所示的方法进行至步骤306,生成电子文档的基础结构图。上文中结合图2已经介绍了如何提炼电子文档的结构化信息。步骤306中即是将所提炼的电子文档的结构化信息以图形的方式展示出来,即结构化的基础结构图。基础结构图的具体形式取决于电子文档的结构组织形式,例如,可能是树形图、索引目录图、层级关系图等。无论是何种样式、效果的图,只要其反映了电子文档自身的结构化信息,即属于步骤306中所指的“基础结构图”。根据本发明的一个实施例,“基础结构图”可以是对文档的部分内容的“基础结构图”而并非整篇文档的全部“基础结构图”。

[0038] 在步骤307中,设定代表不同含义的图例标记。设定图例标记的目的是为了将步骤304或305中的分析结果以图形标记的方式进行表示。例如,可以用矩形标记表示“用户最感兴趣的段落”、可以用椭圆形标记表示“用户最近阅读的段落”、可以用矩形标记表示“最频繁被引用的段落”、可以用单向箭头表示“单向引用”、可以用双向箭头表示“双向引用”、可以用“?”表示“需要跟进”等。本领域技术人员应当理解,可以根据需要设定不同的图例标记,也可以随时修改已经设定的图例标记。

[0039] 接下来在步骤308中,利用在步骤304和305中的分析结果,将图例标记附加至电子文档的基础结构图中,从而生成结构化的文档导视图。例如,如果用户最后阅读的段落是电子文档的第2段,则将表示“用户最近阅读的段落”的椭圆形图例标记添加至电子文档的基础结构图中的代表第2段的相应位置。又例如,假设在步骤304根据规则“阅读次数最多的段落即为用户最感兴趣的段落”分析得到用户对第2章最感兴趣,则将表示“用户最感兴趣的段落”的矩形图例标记添加至电子文档的基础结构图中的代表第2章的相应位置。根据本发明的一个实施例,除了直接将图例标记附加至电子文档的基础结构图中外,为了提供更好的用户体验,还可以在结构化文档导视图加入相应图例标记的属性,例如对于“单向箭头”这一表示单向引用关系的图例标记,可以在结构化文档导视图中的相应的单向箭头上加入属性“关键词”,以使用户能够更直观地了解此单向引用关系是基于何种关键词建立的。根据本发明的一个实施例,当“基础结构图”是文档的部分内容的“基础结构图”的情况下,所生成的结构化文档导视图也是基于文档的部分内容的导视图。例如,可以按照用户的阅读进度,一步步地为用户已经阅读过的文档部分生成结构化的文档导视图。当然,也可以无论用户已经阅读了文档的哪些内容,而直接生成整篇文档全部“基础结构图”,并进而在完整的基础结构图上附加相应的图例标记,以生成结构化文档导视图。

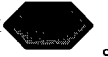
[0040] 需要指出的是,图3所示的方法中的步骤306和307既可以如图所示在完成304和305的分析步骤之后执行,也可以在执行301接收用户操作的步骤之前进行,甚至可以在任何时间段进行,只要在最终生成结构化的文档导视图(步骤308)之前,生成了文档基础结构图并设定好图例标记即可,步骤306和307的执行与否并不影响步骤301至305的执

行。此外，步骤 306 和步骤 307 本身也没有执行顺序先后之分，既可以先设定图例标记再生成电子文档的基础结构图，也可以先生成电子文档的基础结构图再设定图例标记。

[0041] 还需要指出的是，步骤 302 判断后的两条分支（一条分支为步骤 303 和 304，另一分支为步骤 305）并非都是必需的。本领域技术人员在理解了本发明的内容之后完全可以只采用其中一条分支的步骤进行分析，例如只对用户输入的注释进行分析或者只对用户的阅读行为数据进行分析，同样可以在不脱离本发明技术构思的情况下生成结构化文档导视图，同样落入本发明的保护范围。因此，在只对用户输入的注释进行分析或者只对用户的阅读行为数据进行分析时，判断步骤 302 也可以省略。

[0042] 图 4 示出了根据本发明一实施例的所生成的结构化文档导视图的实例。图 4 的右下角示出了图例标记的具体含义。本领域技术人员应当理解，最终生成并展示的结构化文档导视图中并非一定要展示图例标记的具体含义，用户可以另行查找图例标记的定义。图 4 示出的结构化文档导视图中包含 5 种图例标记，分别是：矩形、椭圆形、五边形、问号、单向箭头和双向箭头。除了这 5 种图例标记外，剩余的图形部分即来自于电子文档的基础结构图——展示了电子文档的标题、章节、章节下的段落等结构化信息。下面以问号、五边形和单向箭头三种图例标记为例来具体说明。

[0043] 对于问号（？）这一图例标记而言，如果用户在阅读过程中曾在某处用问号做过标记，在生成导视图时，就会在该段落旁标记一个问号，并且把用户标注的问题作为标签属性。在图 4 中，用户在 2.2 节下面的第 k 段输入了用户注释“？”并又进一步输入了具体的问题，那么在图 4 中第 k 段的位置即会显示这个“？”。根据本发明的一个实施例，用户所输入的具体的问题可以与“？”一起直接显示在结构化文档导视图中。根据本发明的另一实施例，响应于用户点击图 4 中的“？”，进一步显示用户之前所输入的具体问题。

[0044] 对于五边形这一图例标记而言，如果经过分析得知，用户对某个段落最感兴趣，那么在生成导视图时，会在该段落旁附加一个代表用户最感兴趣的标签 。并且会把分析得出的其他额外信息，如用户最感兴趣的行、关键字，以及访问频率等作为标签属性。从图 4 中可以看出，在这篇电子文档中，用户对第 3 章第 t 段最感兴趣。图 4 中也并未直接显示这一最感兴趣的段落，而是响应于用户点击了第 3 章的图标，才最终显示第 3 章下面的更具体的结构化导视图。当然，本领域技术人员也完全可以令所有结构化导视图信息同时显示。

[0045] 对于单向箭头这一图例标记而言，如果经分析得知，用户认为段落 a 与段落 h 之间有逻辑引用关系，并且关键词 X 被视为重要的链接点。那么在生成导视图时，就会添加从段落 h 指向 a 的箭头标注，且在箭线上注明关键字 X。该箭头包括的属性包括引用次数，关键字等。

[0046] 以上是生成基于用户理解和兴趣的文档导向图的基本流程。图的绘制可以利用现有的控件技术实现。根据本发明的一个实施例，结构化文档导视图是可编辑的。也就是说，导视图生成后，用户可以对自动生成的文档导视图上的任一节点和图元作相应地增，删，改。也可以将该文档或书籍的导视图导出 pdf 或 xml 等格式，与其他读者共享。结构化文档导视图是可编辑的也可体现为：根据该文档的其它用户对该文档的操作而进一步完善、修正已经存在的结构化文档导视图，从而使得所生成的结构化文档导视图能够体现多个用户对该文档的阅读历史、兴趣、习惯。

[0047] 图 5 示出了根据本发明一实施例的用于生成结构化文档导视图的系统的框图。图 5 所示的系统在整体上由系统 500 表示。具体地,系统 500 包括接收装置 501,被配置为接收用户对所述文档的操作;分析装置 502,被配置为对所接收到的操作进行分析;以及导视图生成装置 503,被配置为利用所述分析结果和所述文档的基础结构,生成结构化文档导视图。本领域技术人员应当理解,系统 500 中的装置 501-503 分别对应于图 2 所示的方法中的步骤 201-203,在此不再赘述。

[0048] 附图中的流程图和框图,图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分,所述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个接连地表示的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意,框图和 / 或流程图中的每个方框、以及框图和 / 或流程图中的方框的组合,可以用执行规定的功能或操作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0049] 虽然以上结合具体实例,对本发明的利用远程应用处理本地文件的系统及方法进行了详细描述,但本发明并不限于此。本领域普通技术人员能够在说明书教导之下对本发明进行多种变换、替换和修改而不偏离本发明的精神和范围。应该理解,所有这样的变化、替换、修改仍然落入本发明的保护范围之内。本发明的保护范围由所附权利要求来限定。

计算系统 100

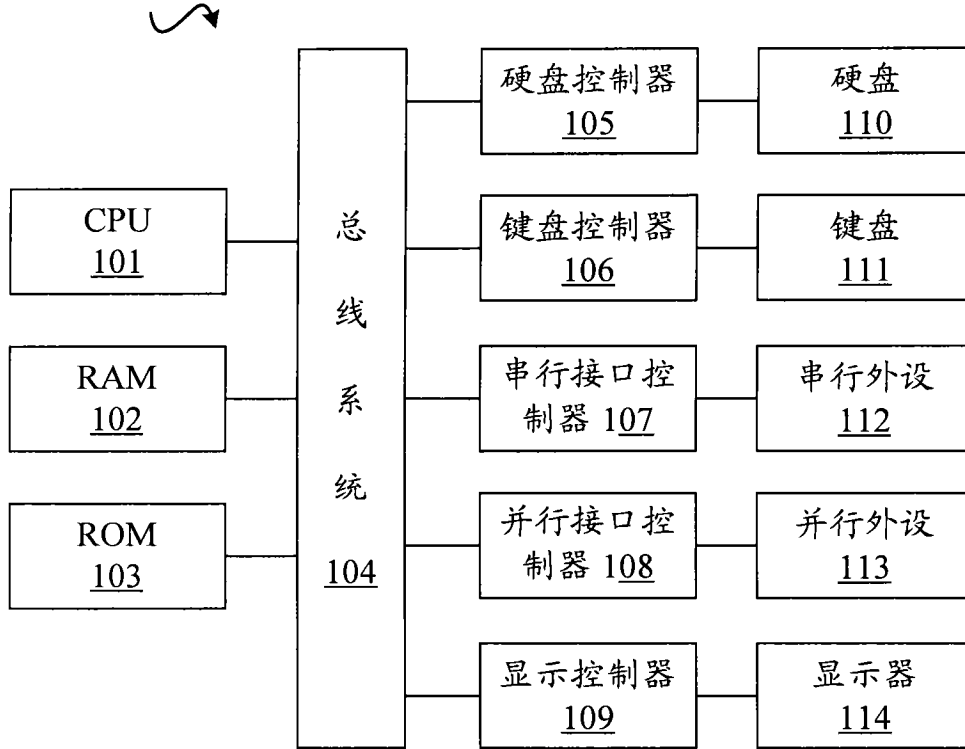


图 1

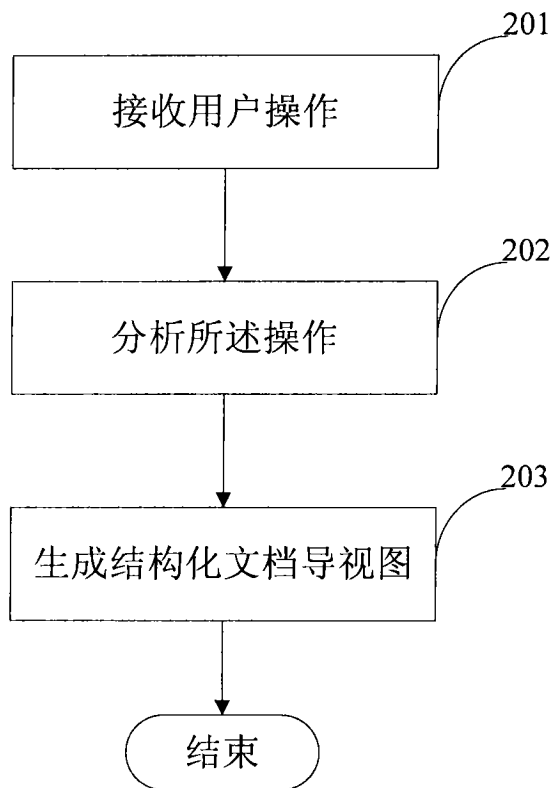


图 2

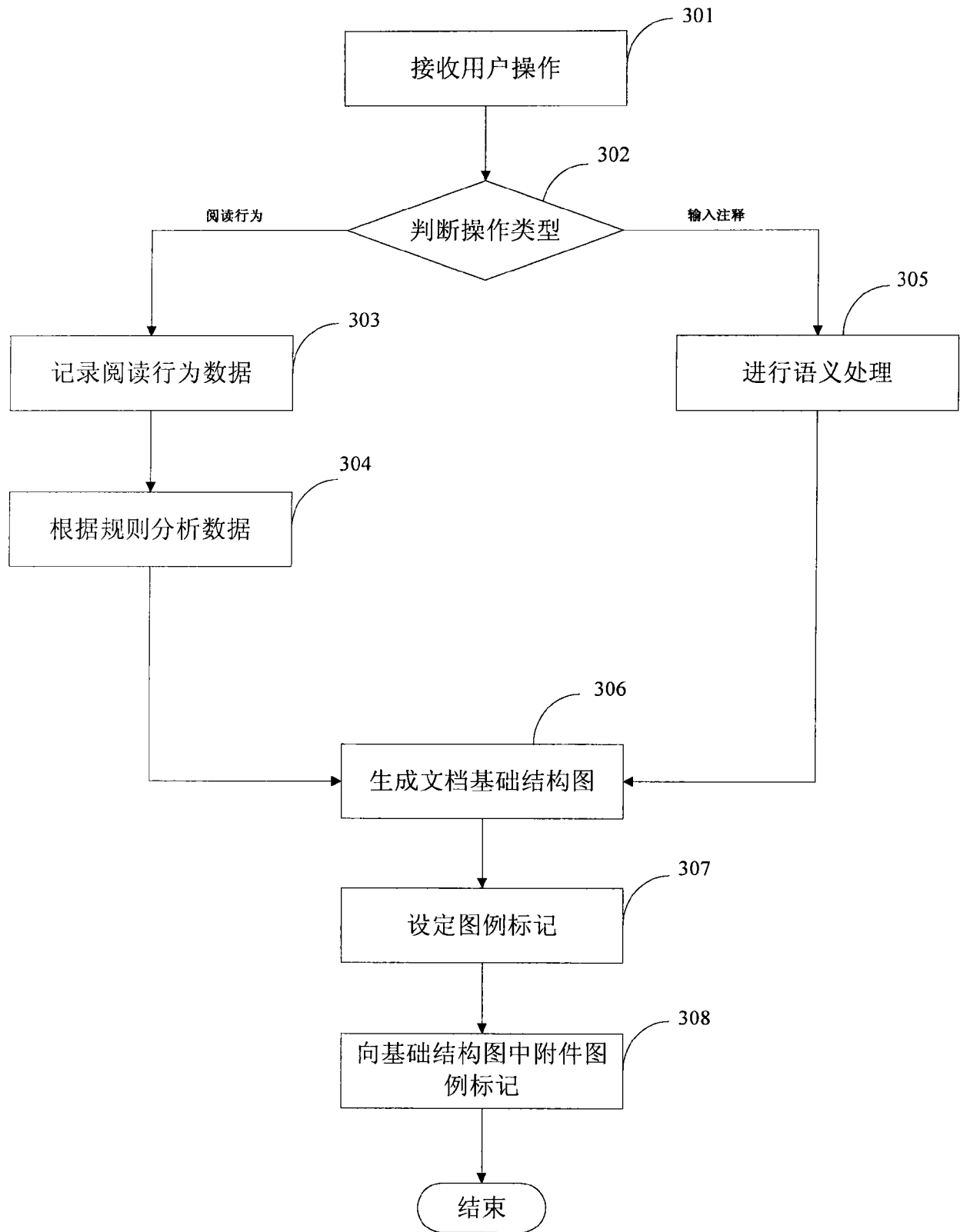


图 3

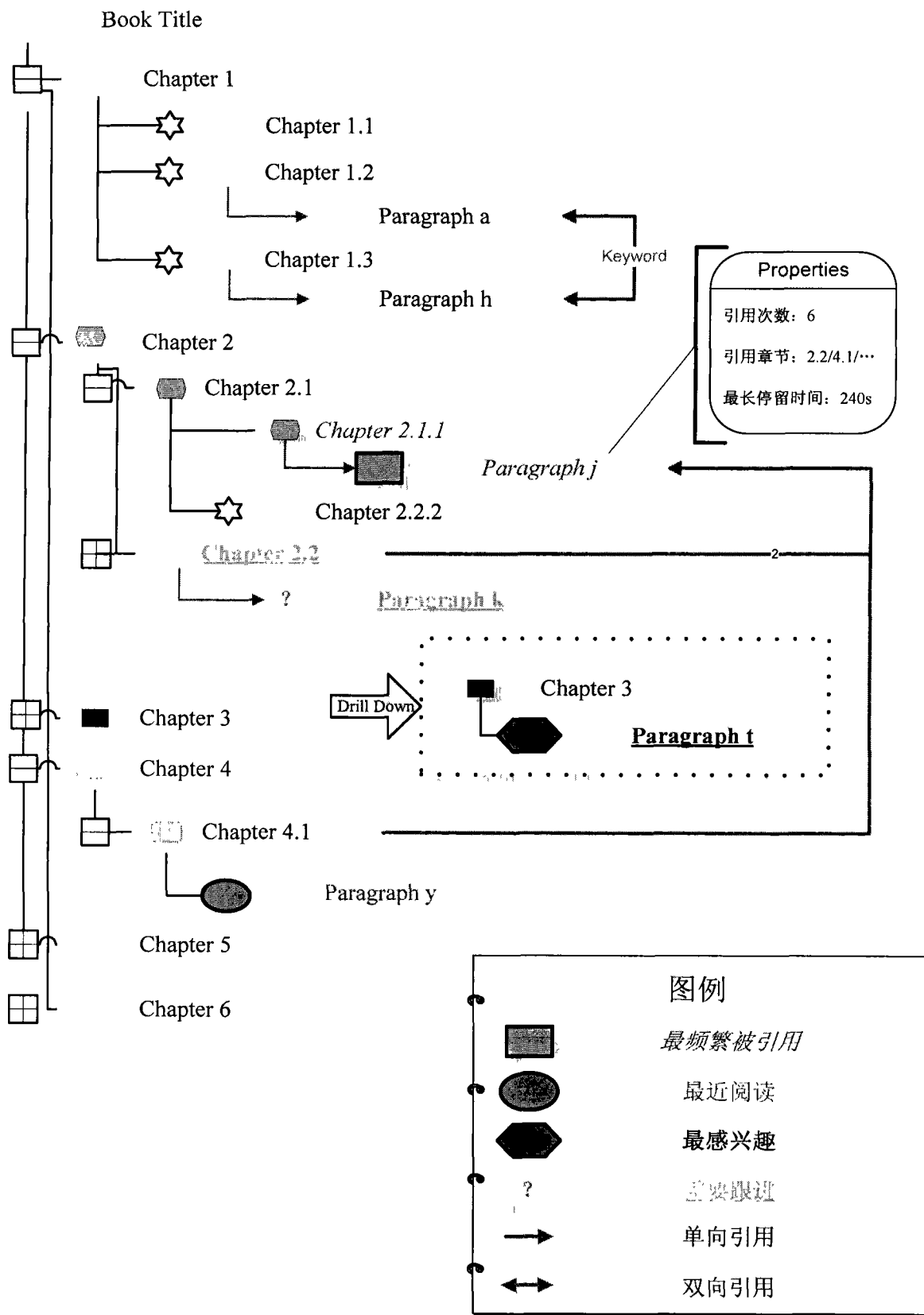


图 4

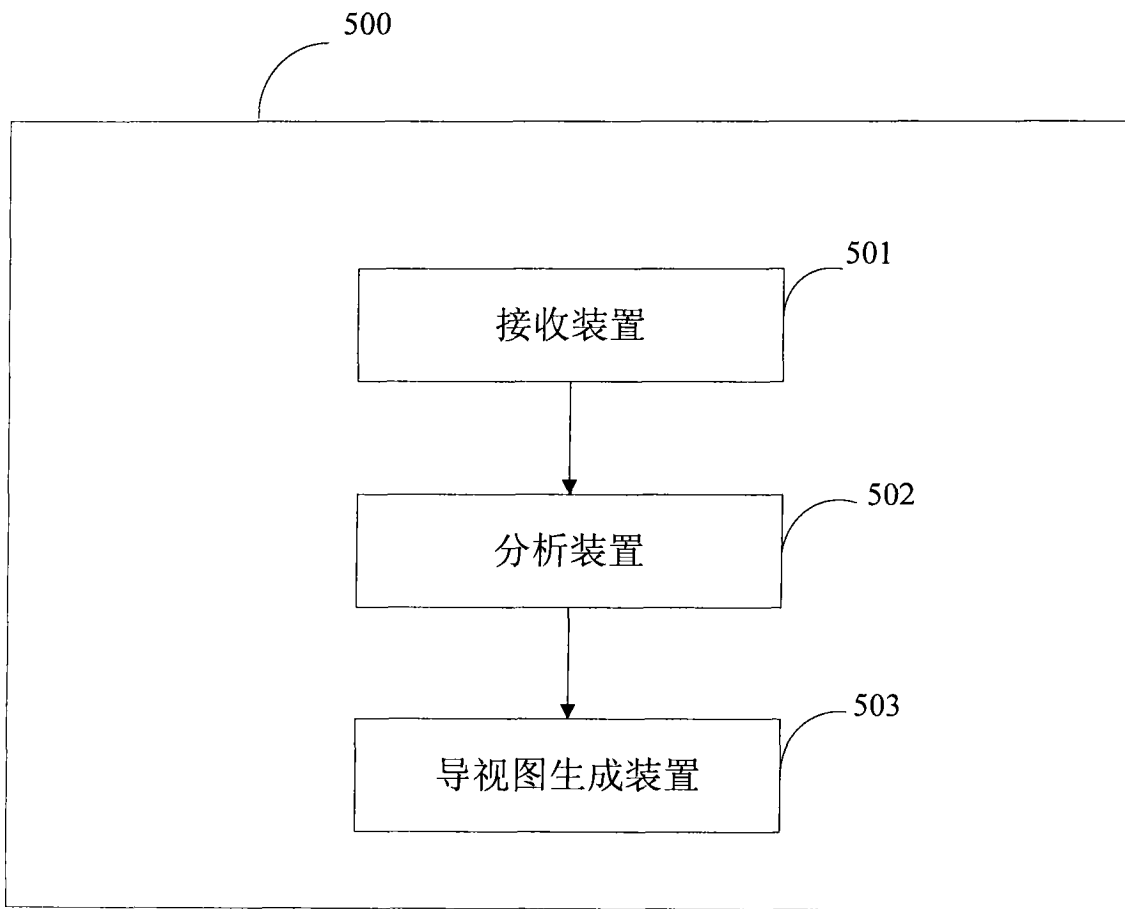


图 5